Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations

Article

Published Version

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD025480

Key Points:
- Changes in climate extremes over the twentieth century are analyzed in gridded observations, reanalyses, and SST-driven model runs
- On global average there is consistent warming of temperature and intensification of wet precipitation extremes since the midtwentieth century
- There are larger interdata set uncertainties in the first half of the twentieth century

Supporting Information:
- Supporting Information S1

Correspondence to:
M. G. Donat,
m.donat@unsw.edu.au

Citation:
Donat, M. G., L. V. Alexander, N. Herold, and A. J. Dittus (2016), Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations, *J. Geophys. Res. Atmos.*, *121*, 11,174–11,189, doi:10.1002/2016JD025480.

Received 8 JUN 2016
Accepted 1 SEP 2016
Accepted article online 14 SEP 2016
Published online 1 OCT 2016

Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations

Markus G. Donat[1], Lisa V. Alexander[1], Nicholas Herold[1], and Andrea J. Dittus[2,3]

[1]Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia, [2]School of Earth Sciences and ARC Centre of Excellence for Climate System Science, University of Melbourne, Melbourne, Victoria, Australia, [3]Now at School of Earth, Atmosphere and Environment, Monash University, Clayton, Victoria, Australia

Abstract Knowledge about long-term changes in climate extremes is vital to better understand multidecadal climate variability and long-term changes and to place today's extreme events in a historical context. While global changes in temperature and precipitation extremes since the midtwentieth century are well studied, knowledge about century-scale changes is limited. This paper analyses a range of largely independent observations-based data sets covering 1901–2010 for long-term changes and interannual variability in daily scale temperature and precipitation extremes. We compare across data sets for consistency to ascertain our confidence in century-scale changes in extremes. We find consistent warming trends in temperature extremes globally and in most land areas over the past century. For precipitation extremes we find global tendencies toward more intense rainfall throughout much of the twentieth century; however, local changes are spatially more variable. While global time series of the different data sets agree well after about 1950, they often show different changes during the first half of the twentieth century. In regions with good observational coverage, gridded observations and reanalyses agree well throughout the entire past century. Simulations with an atmospheric model suggest that ocean temperatures and sea ice may explain up to about 50% of interannual variability in the global average of temperature extremes, and about 15% in the global average of moderate precipitation extremes, but local correlations are mostly significant only in low latitudes.

1. Introduction

Climate extremes often have severe impacts on different aspects of society, infrastructure, and ecosystems [*Intergovernmental Panel on Climate Change*, 2012]. In a changing climate, it is important to know how characteristics of extremes are changing, and will change in the future, to enable appropriate adaptation. Relatively good observational data coverage, suitable to assess past changes in temperature and precipitation extremes, exists over land since the midtwentieth century [*Alexander et al*., 2006; *Donat et al*., 2013a, 2013b]. These observations show widespread warming trends in temperature extremes, while changes in precipitation extremes are spatially more heterogeneous.

However, going back in time prior to the midtwentieth century, reliable observations become sparse, and little work has been done to assess changes in extremes on century time scales [*Donat et al*., 2013b]. The knowledge of changes in extremes on longer time scales is important to understand decadal-scale variability [e.g., *Salinger*, 2005; *King et al*., 2013], to consider recent extremes in a historical context [e.g., *Donat et al*., 2013a], and also to evaluate climate models [e.g., *Sillmann et al*., 2013].

Several observations-based global data sets have recently become available that allow the investigation of extremes over the past century. One is based on interpolated in situ observations (HadEX2) [*Donat et al*., 2013b]. Others are reanalyses: dynamical atmospheric models that were constrained by observations of the surface pressure, sea surface temperatures, and sea ice concentrations, in addition to increasing greenhouse gas concentrations. These are the Twentieth Century Reanalysis by the National Oceanic and Atmospheric Administration, hereafter 20CR [*Compo et al*., 2011], and the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of the twentieth century [*Poli et al*., 2016], hereafter ERA-20C. In addition, ECMWF also provides an ensemble of simulations with the same atmospheric model used for ERA-20C, only driven with monthly ocean surface temperatures and sea ice concentrations,

but without assimilating variables related to atmospheric flow. This ensemble of ocean-driven simulations is called ERA-20CM [*Hersbach et al*., 2015]. Note that both 20CR and ERA-20C use similar observational constraints from historical pressure records but different atmospheric models and assimilation systems and are therefore only partly independent realizations of the past climate. Both data sets are, however, independent from the HadEX2 gridded observations that are based on in situ temperature and precipitation records.

A comprehensive comparison of climate extremes across several gridded observations and operational reanalyses has shown that most state-of-the-art reanalyses are consistent with in situ-based data sets for the most recent three decades when satellite data were available for assimilation into the atmospheric models [*Donat et al*., 2014]. Good agreement between the ECMWF ERA-Interim reanalysis and observations during the most recent three decades was also found in a regional evaluation of extreme temperatures specifically over Europe [*Cornes and Jones*, 2013]. However, it is questionable if reanalyses are suitable for investigations of long-term changes due to inhomogeneities in their input data [*Thorne and Vose*, 2010]. Regarding annual average temperatures on a century-long time scale, the 20CR has previously been used to independently confirm warming in global data sets of temperature observations over the 1901–2010 period [*Compo et al*., 2013]. However, climate extremes often show a different behavior compared to average conditions [*Seneviratne et al*., 2014], and it remains to be investigated how robust different estimates of extreme climate conditions are between the independent long-term data sets.

The 20CR has previously been used to investigate changes in extremes but mostly on regional scales. For example, long-term upward trends in European storminess have been found [*Donat et al*., 2011] but were argued to be, at least in part, artifacts related to increasing density in the observational network used for the assimilation of atmospheric flow [*Krueger et al*., 2013]. However, *Wang et al*. [2013] highlight consistency in extreme pressure gradients between 20CR and observations over the NE Atlantic, and despite large inhomogeneities in the ensemble mean pressure fields, *Pepler et al*. [2016] showed that the full ensemble could be used for a long-term assessment of midlatitude cyclone frequency on the east coast of Australia. For temperature extremes over North America, there is reasonably good agreement between 20CR and gridded station observations throughout the entire twentieth century [*Donat et al*., 2016]. Temperature and precipitation extremes on a global scale, however, have not been previously compared between century-long reanalyses and data sets based on direct observations of temperature and precipitation.

In this study, we compare indices of extreme temperature and precipitation calculated from two century-long reanalyses, 20CR and ERA-20C, with gridded extremes from the HadEX2 data set over the period 1901–2010. Our aims are to (i) provide an assessment of century-scale changes and variability for different measures of temperature and precipitation extremes. Further, based on the agreement between these independent data sets, we will (ii) infer the robustness of our knowledge about changes in climate extremes over the past century. By comparing observational extremes to atmospheric model simulations driven by ocean fields, we (iii) also estimate to what extent interannual variability of extremes may be driven by ocean surface temperatures and sea ice.

2. Data and Methods

In this study we consider a subset of the 27 indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI) to investigate different aspects of temperature and precipitation extremes [*Zhang et al*., 2011]. These indices are calculated from daily maximum and minimum temperatures and total daily precipitation amounts. Table 1 shows an overview of the indices that are discussed in this study.

HadEX2 provides global land-based gridded fields of the ETCCDI indices [*Donat et al*., 2013b]. It is based on a network of several thousand high-quality meteorological stations. The annual extremes are first calculated at each station location before the indices are interpolated on a global 3.75° × 2.5° longitude-latitude grid. HadEX2 covers the period 1901 to 2010, but the spatial coverage is variable in time [*Donat et al*., 2013b].

20CR version 2 is a global reanalysis that assimilates hourly and synoptic barometric pressure observations, monthly averaged sea surface temperature (SST), and sea ice concentration (SIC) fields into an atmospheric circulation model, in addition to prescribing observed radiative forcings [*Compo et al*., 2011]. To account for uncertainties in the sparse observational input data, an Ensemble Kalman Filter technique is used to produce

Table 1. Climate Extreme Indices Analyzed in This Study

ID	Index Name	Index Definition	Unit
TXx	Hottest day	Annual maximum value of daily maximum temperature	°C
TNx	Warmest night	Annual maximum value of daily minimum temperature	°C
TXn	Coldest day	Annual minimum value of daily maximum temperature	°C
TNn	Coldest night	Annual minimum value of daily minimum temperature	°C
TN10p	Cool nights	Annual count when daily minimum temperature $<$ 10th percentile	days
TX10p	Cool days	Annual count when daily maximum temperature $<$ 10th percentile	days
TN90p	Warm nights	Annual count when daily minimum temperature $>$ 90th percentile	days
TX90p	Warm days	Annual count when daily maximum temperature $>$ 90th percentile	days
FD	Frost days	Annual count when daily minimum temperature $<$ 0°C	days
SU	Summer days	Annual count when daily maximum temperature $>$ 25°C	days
Rx1day	Maximum 1 day precipitation amount	Annual maximum 1 day precipitation	mm
R10mm	Number of heavy precipitation days	Annual count when daily precipitation $\geq$ 10 mm	days
R95p	Precipitation from very wet days	Annual total precipitation from days $>$ 95th percentile	mm
CDD	Consecutive dry days	Maximum number of consecutive days when precipitation $<$ 1 mm	days

an ensemble of 56 realizations. The 20CRv2 provides 1.875° × 1.9° horizontal resolution global gridded fields for the period 1871 to 2010. The ETCCDI indices are calculated from daily gridded fields of daily maximum and minimum temperature and daily precipitation amounts for each of the 56 ensemble members.

The ERA-20C is another global reanalysis that was produced by assimilating surface pressure and marine wind observations into a numerical model of the atmosphere [*Poli et al*., 2016] and using SST and SIC fields as boundary conditions. The ERA-20C was produced with the ECMWF Integrated Forecasting System (IFS) version Cy38r1, at a horizontal resolution of approximately 1.25° (spectral truncation T159). Output fields were downloaded from the ECMWF data server on a regular 1.5° × 1.5° grid. The ERA-20C covers the period 1900 to 2010. Global fields of daily maximum temperatures were calculated as the daily maximum of 6-hourly instantaneous fields and daily minimum fields accordingly as the minimum of the 6-hourly values. Daily precipitation totals from ERA-20C were provided as 24 h cumulates as a forecast variable. Unlike 20CR that provided a relatively large ensemble to account for uncertainties in the observational fields, the final version of ERA-20C was produced only as a single realization.

In addition to the ERA-20C reanalysis, ECMWF also provides an ensemble of simulations with the same IFS model version Cy38r1, driven only by observational SST and SIC fields, but without assimilating pressure and wind observations [*Hersbach et al*., 2015]. Hence, this ensemble follows a setup as used in the Atmospheric Model Intercomparison Project (AMIP) [*Gates*, 1992]. ERA-20CM consists of 10 ensemble members that were produced using different realizations of SST and SIC fields from the HadISST2 data set [*Titchner and Rayner*, 2014]. Temperatures from ERA-20CM are 3-hourly instantaneous fields, and total precipitation comes as 3-hourly accumulated values. Hence, daily maximum temperatures were calculated as the daily maximum of the 3-hourly values and daily minimum temperatures as the minimum of 3-hourly values accordingly. Daily precipitation totals were calculated as daily sums of the 3-hourly totals. We will hereafter use the terms AMIP-style runs or AMIP ensemble to refer to the results based on ERA-20CM.

The ETCCDI climate extreme indices were calculated on the native grids of 20CR, ERA-20C, and ERA-20CM. Prior to the analyses, however, the extreme indices from all data sets were regridded to the HadEX2 grid which is the one with the coarsest resolution across all data sets used here. Regridding is necessary to mask all data sets to the same spatial coverage when calculating global and regional average time series, for example. Also, for calculating local correlations between data sets, the same grid was necessary.

Note that the order of operation to obtain the fields of annual extremes is different between the data sets. While HadEX2 first calculates the extreme indices before gridding them, reanalyses and climate model extremes are calculated from gridded daily fields. These different orders of operation can cause systematic differences in the absolute values of some extreme measures [*Donat et al*., 2014; *Avila et al*., 2015]. Extremes calculated from daily grids, such as climate model and reanalysis data, are usually "less extreme" compared to grids calculated from station extremes; i.e., lower absolute maximum values and higher minimum values are expected in extremes from daily grids when compared to grids of station extremes. This mainly affects indices based on absolute values, while indices based on relative thresholds (e.g., percentile

exceedance) generally are more robust across the different gridded data sets if calculated relative to their own percentiles [*Sillmann et al*., 2014]. Also, while the order of operation issues have effects on the actual values of extremes, interannual variability and long-term changes are generally less sensitive to these differences in calculating the global grids of extremes [*Donat et al*., 2014]. Therefore, we use anomalies from the common climatological period 1961–1990 to intercompare the data sets.

We present analyses for the period 1901–2010, which is common to all data sets. Global average time series for each index are calculated as area-weighted averages of local annual anomalies from the 1961–1990 climatological mean. To minimize artifacts from variable spatial coverage that may lead to variability in the global time series, we only use grid cells that have at least 90 years of valid data for the global average calculation. This has the caveat that coverage contributing to the "global" time series is relatively limited, and in reality this global time series mainly is an average of grid cells in North America, Europe, parts of Asia, and Australia. Note that the spatial coverage varies for different indices in HadEX2. This is because the spatial correlation structure of the underlying station data is taken into account when calculating the grids [*Alexander et al*., 2006; *Donat et al*., 2013b]; indices with a larger decorrelation length scale have more grid cells with data compared to indices with a shorter decorrelation length scale.

Difference maps are shown for averages of the climate extreme indices during the last 20 years (1991–2010) relative to the first 20 years (1901–1920) and a period in the middle of the twentieth century (1951–1970), when data already have a reasonable global coverage and before strongest warming occurred [see, e.g., *Donat et al*., 2013b].

We calculate Spearman rank correlations of global average and regional grid cell time series between the different data sets to assess their agreement in the interannual variability of extremes. All time series were detrended before calculating the correlations.

3. Results

We discuss results for some selected indices that represent extremes of temperature (cold and warm) and precipitation (wet and dry). In order to avoid a large number of very similar figures for different indices, we restrict ourselves to only showing one or two indices per subsection. Results from additional indices discussed are available as supporting information figures. Each of the following subsections contains a figure of changes in an index representative for the frequency of extremes and a figure of an index representing the intensity of extremes, respectively. Just for dry extremes there is only one index available in the suite of ETCCDI indices, CDD, and this is a measure of duration.

3.1. Warm Extremes

Gridded observations (HadEX2) show that the frequency of warm days (TX90p) has increased over the past century in most regions of the globe (Figure 1), with the exceptions of the warming hole over eastern North America [*Portmann et al*., 2009; *Misra et al*., 2012] and the southern and western parts of South America [*Skansi et al*., 2013]. Both reanalyses also seem to show some realization of a warming hole over central or eastern North America. The 20CR, however, shows different patterns of change in several regions, particularly in northern Eurasia, large parts of South America, and high latitudes of North America, where no increases or even decreases of TX90p are found. The ERA-20C shows a reasonably similar spatial pattern of change to HadEX2 (spatial correlation about 0.7), but warming is stronger than observed over large parts of Eurasia. However, a small region of cooling in northeastern Asia when comparing the earliest and latest 20 year time slices indicates that ERA-20C simulates a large number of warm days in this region in the early twentieth century. The AMIP ensemble shows warming in all land regions, and the global patterns of change are more similar to the gridded observations than was the case for the two reanalyses (spatial correlation about 0.8 for this index).

The global average time series (using only grid cells that are 90% complete in HadEX2) show significant warming in all data sets since the middle of the twentieth century, with strongest warming occurring after about 1975. There is, however, some spread between the different data sets. The 20CR shows the smallest increase in warm days (about 11 days more in the early 2000s relative to the 1961–1990 climatology), and ERA-20C shows the strongest increase (about +20 days).

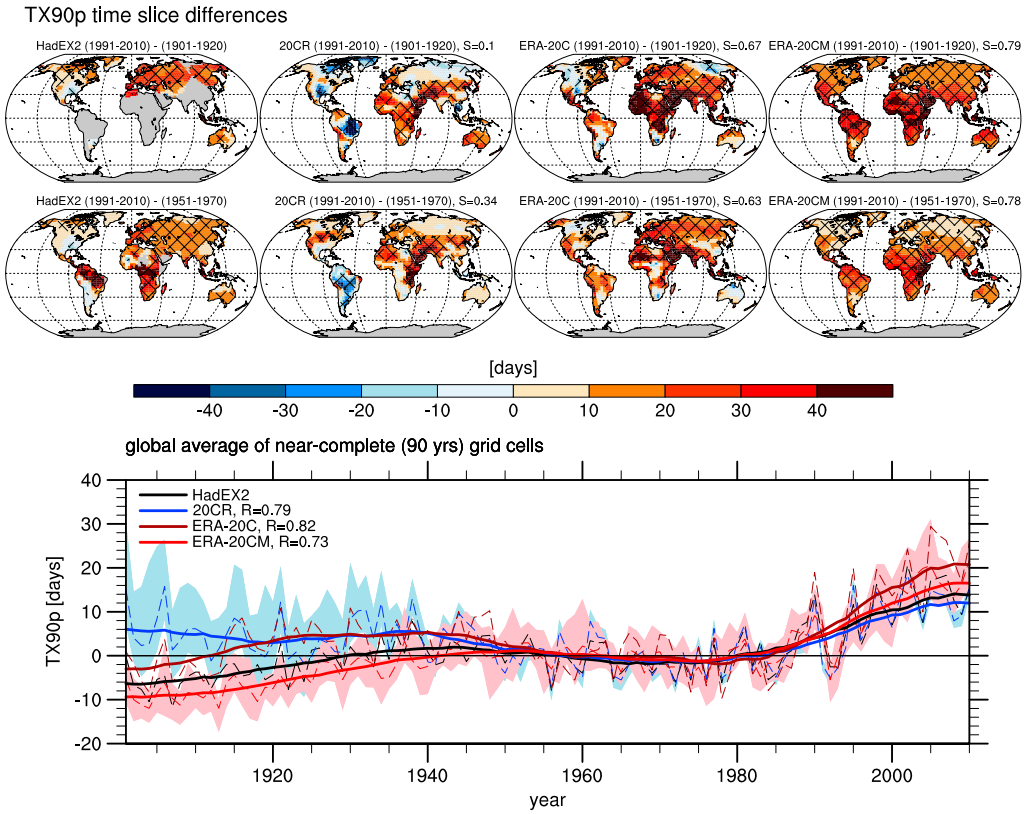

Figure 1. Regional and global changes in the number of warm days (TX90p). (first row) Time slice differences between the most recent 20 year average (1991–2010) and the earliest 20 year average (1901–1920) of the analysis period. (second row) Time slice differences between the most recent 20 year average (1991–2010) and a 20 year period in the midtwentieth century (1951–1970). Difference maps are shown for (from left to right) HadEX2, 20CR ensemble mean, ERA-20C, and ERA-20CM ensemble mean. Hatching indicates where local changes are significantly different from zero (Student's *t* test, $p \leq 0.05$). *S* indicates the uncentered spatial correlation of the change patterns in 20CR, ERA-20C, and ERA-20CM with the change pattern in HadEX2, only using grid cells for which HadEX2 changes could be calculated. (bottom) Global average time series of the four data sets, only using grid cells that are near complete (90 years or more with valid data) in HadEX2. Thin dashed lines show the annual values and bold solid lines the 21-point Gaussian filtered data. Blue and red shaded areas represent the ensemble spread for 20CR and ERA-20CM, respectively. *R* indicates the temporal correlations of the different detrended annual time series with detrended HadEX2.

The different global average time series show good agreement back to around 1940 but diverge for earlier decades. In particular, the two reanalyses that assimilate pressure observations (and thus large-scale atmospheric flow) are warmer during the early twentieth century compared to gridded observations and the AMIP simulations. This is mainly related to warmer conditions during early decades in high northern latitudes, where maps of time slice differences (Figure 1, first row) did not show warming.

Temporal correlations between the different data sets are high throughout the 110 year investigation period. Detrended global average time series from the two reanalyses show correlations of about 0.8 with HadEX2 (20CR: $R = 0.79$, ERA-20C: $R = 0.82$). And even the AMIP ensemble mean shows a correlation of about 0.73 with HadEX2, suggesting that interannual variability of warm extremes on a global scale is to a large extent driven by ocean surface temperatures.

Local correlations between the reanalyses and gridded observations are highest ($R > 0.7$) in regions where the observational network is dense and of relatively high quality, i.e., over North America, Eurasia, and Australia (Figure 2). In contrast, the AMIP ensemble has low correlations ($R < 0.4$) with gridded observations in these extratropical regions but moderate to high correlations ($R > 0.6$) in some low-latitude areas, particularly South America and Africa. This shows that SST variability has the strongest influence on local temperature extremes in tropical and subtropical regions.

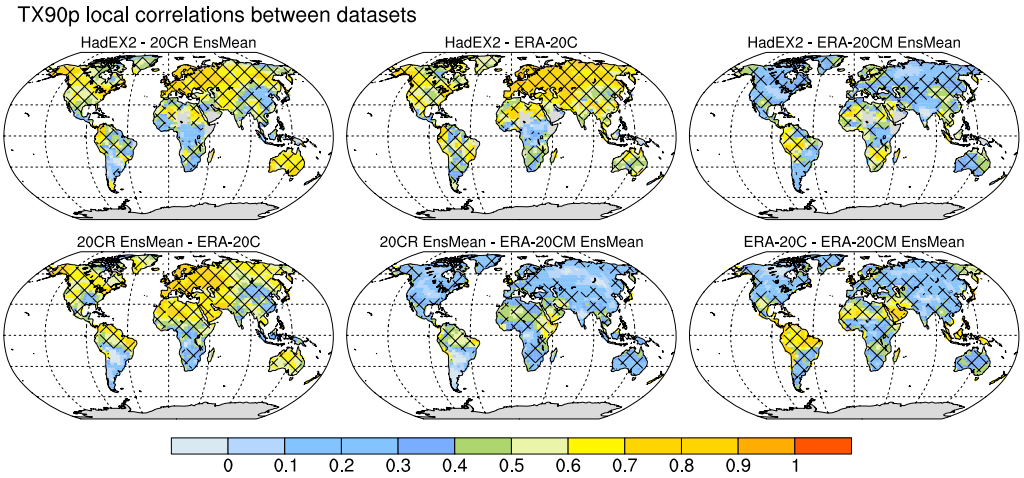

Figure 2. Maps of local correlations between the different data sets for TX90p. The time series from each grid cell were detrended prior to calculating correlations. Hatching indicates where correlations are significant ($p \leq 0.05$).

TX90p is a measure of rather moderate temperature extremes, as by definition about 36.5 days exceed the 90th percentile on average per year. We also consider a more "extreme" measure, the temperature of the hottest day of each year (TXx). Note that TX90p counts warm days throughout the entire year, whereas TXx always occurs during the warm season. Difference patterns appear reasonably similar to changes in TX90p, although the local cooling in 20CR between the early twentieth century and the most recent 20 year period

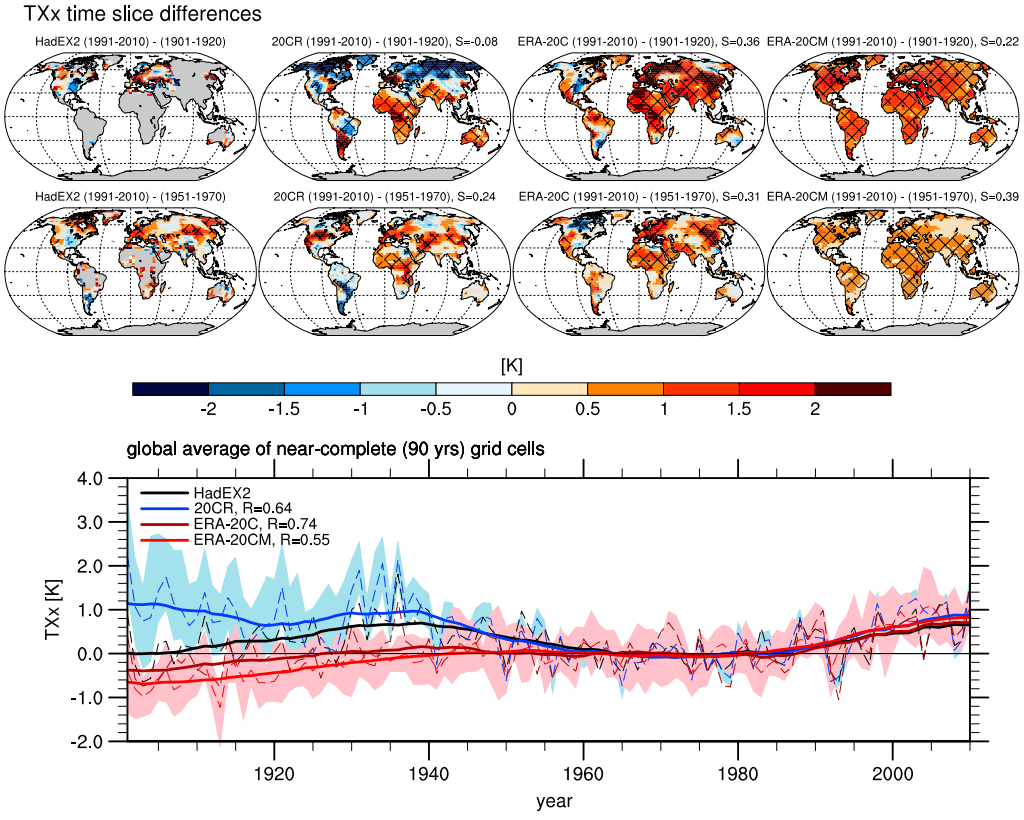

Figure 3. Regional and global changes in the hottest day of the year (TXx). For detailed description of the plot see Figure 1 caption.

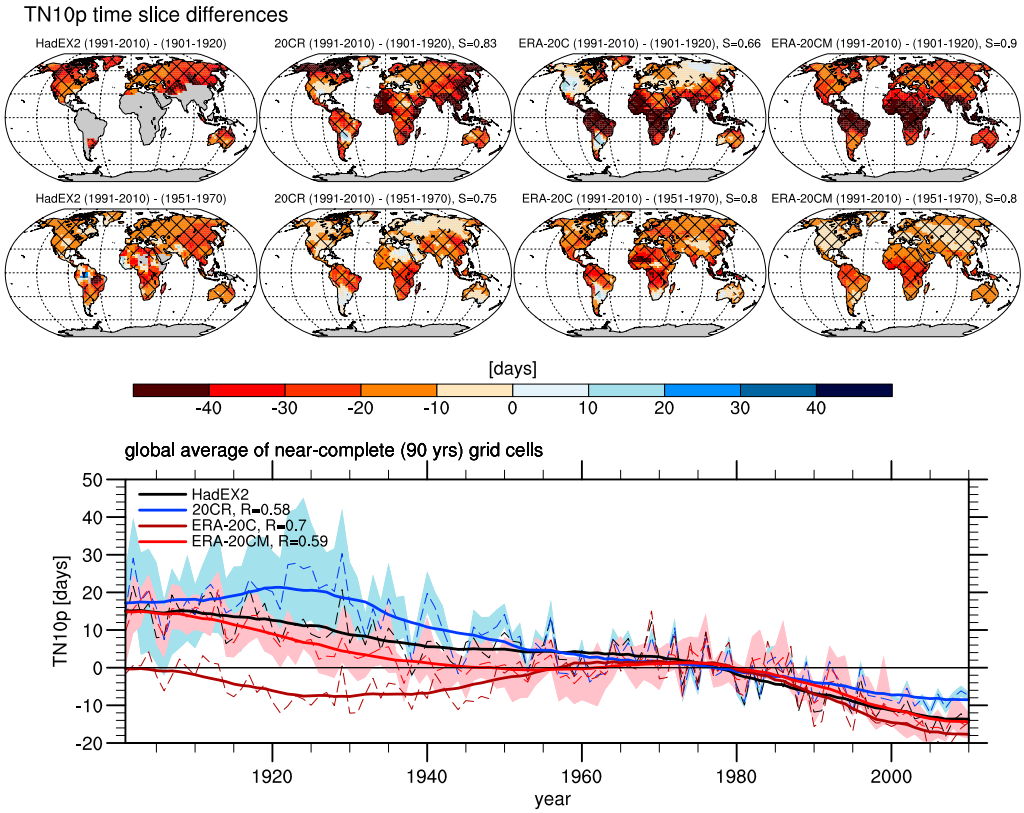

Figure 4. Regional and global changes in the number of cool nights (TN10p). For detailed description of the plot see Figure 1 caption.

in high northern latitudes is even more pronounced for TXx (Figure 3). This suggests that the northern high-latitude warm bias in 20CR in the early twentieth century is particularly strong during summer. The gridded observations and reanalyses all show some areas where TXx did not increase over the past century (e.g., the warming hole over the U.S. and some parts of South America). In contrast, the AMIP ensemble shows TXx increases over all land areas. Spatial correlations with the gridded observations are substantially lower than for TX90p.

All four data sets agree that global average TXx has increased since the middle of the twentieth century, but time series show different behavior across the different data sets for the early decades. While ERA-20C and the AMIP ensemble show increases throughout the entire twentieth century, 20CR in particular shows higher global average TXx in the early twentieth century than in the early 2000s. The time slice difference maps (Figure 3, first row) suggest that relatively high TXx in 20CR during the early decades occur mainly in high northern latitudes. The gridded observations (HadEX2) also show high global average TXx in the 1930s. Note, however, that due to sparse observational coverage in early decades, the global time series are dominated by Northern Hemisphere extratropics, which are known to have experienced relatively warm climate conditions and hot summers during the 1930s [*Bengtsson et al.*, 2004; *Drinkwater*, 2006; *Donat et al.*, 2016].

Correlations of the global average time series with the HadEX2 gridded observations are lower than for TX90p but still moderately high: 0.64 for 20CR, 0.74 for ERA-20C, and 0.55 for the AMIP ensemble mean. Note that TXx, defined as a block maximum, is noisier in nature than TX90p that counts the frequency of relatively moderate extremes; therefore, lower correlations are to be expected.

Also, the local correlations between the different TXx data sets are about 0.1 lower than for TX90p (not shown), but spatial patterns are similar in that highest local correlations between the reanalyses and gridded observations are found over North America, Eurasia, and Australia. Local correlations between the AMIP

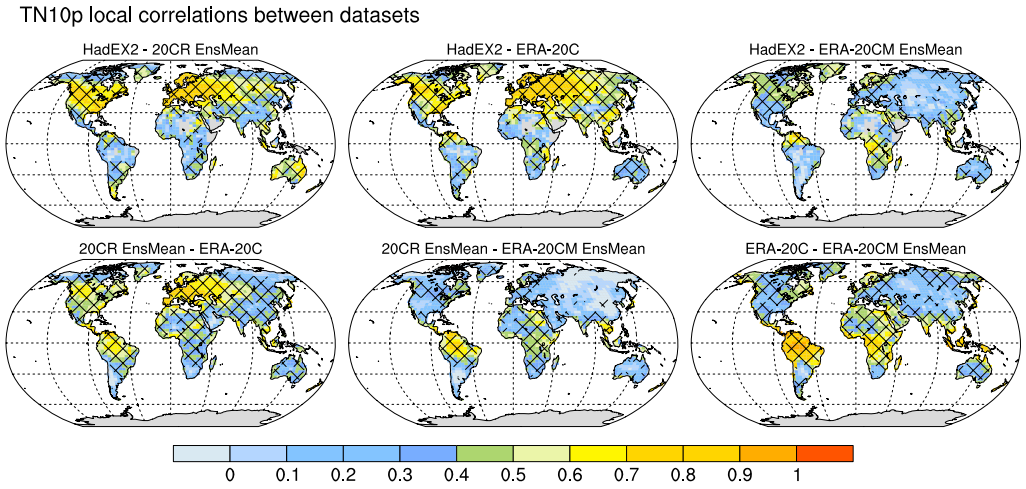

Figure 5. Maps of local correlations between the different data sets for TN10p. The time series from each grid cell were detrended prior to calculating correlations. Hatching indicates where correlations are significant ($p \leq 0.05$).

ensemble mean and the data sets that make use of (some kind of) land-based observations are generally not significant ($p < 0.05$).

Broadly similar characteristics to TX90p and TXx are also found for other measures of warm extremes, such as frequency of warm nights (TN90p), temperature of the warmest night (TNx), or number of summer days (SU), respectively (see supporting information Figures S1–S6).

3.2. Cold Extremes

The frequency of cool nights (TN10p) has decreased, consistent with warming, throughout the past century in virtually all areas with observational coverage (Figure 4). Also, the century-long reanalyses show TN10p decreases in most regions, although there are some regions with no warming signals over North America, South America, and northern Asia. The AMIP ensemble shows changes toward less frequent cool nights over all land areas. Spatial correlations of the change patterns with HadEX2 are moderate to high (ranging from 0.66 to 0.90) for both reanalyses and the ERA-20CM runs.

Also the global average time series show decreases in TN10p over the past century, and all data sets show similar behavior over the most recent 50 years (decreases about 12 to 20 days). Again, the time series show larger differences during the early twentieth century, with the two reanalyses data sets forming the upper and the lower bounds. While 20CR shows on average about 20 more cool nights in the 1920s compared to the 1961–1990 average, ERA-20C has global average TN10p more similar to the 1961–1990 average during the early decades. This is related to relatively warm conditions (for the cool nights index) over North America, southern South America, and northern Asia during the early twentieth century in ERA-20C.

Correlations between the data sets, as a measure of agreement in interannual variability, are moderately high for the global average time series from the two reanalyses (20CR versus HadEX2: $R = 0.58$; ERA-20C versus HadEX2: $R = 0.7$) and also for the AMIP ensemble ($R = 0.59$). As with the warm extremes, local correlations between gridded observations and reanalyses are highest in those extratropical regions with good observational coverage: up to about 0.8 over Europe and North America (Figure 5). Again, the AMIP ensemble shows lower correlations with the other data sets in the extratropics but higher in low latitudes (up to about 0.5 to 0.8, depending on the specific combination of data sets).

The coldest night of the year (TNn), as a measure for a "more extreme" cold extreme, has warmed by more than 2 K over the past century in large parts of Asia and North America (Figure 6). This pattern of strong regional warming is found in the gridded observations and the two century-long reanalyses. The AMIP ensemble shows weaker regional warming in these areas; however, the warming signal is spatially smoother in this data set, with almost all land areas showing warming between 1 and 2 K over the past century.

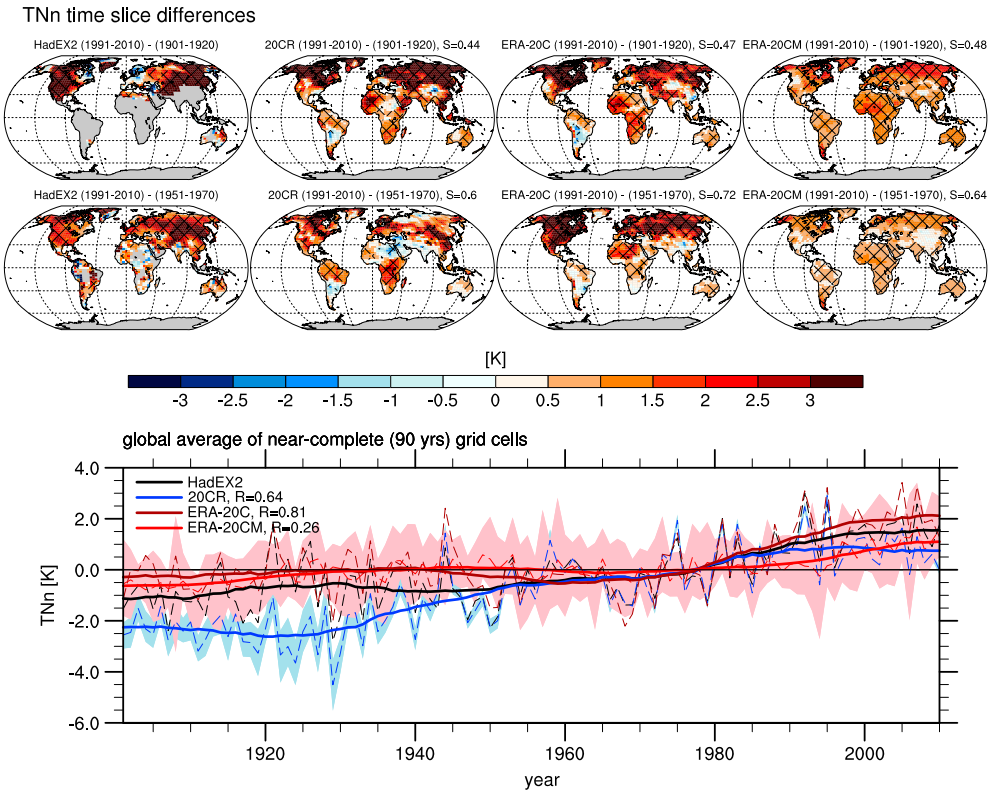

Figure 6. Regional and global changes in the coldest night of the year (TNn). For detailed description of the plot see Figure 1 caption.

On global average, 20CR shows lowest TNn values in the early twentieth century (about 2 K below the 1961–1990 average) and smallest increases during the most recent two decades but has the strongest increase over the entire century (about 3 K). HadEX2 and ERA-20C increase by about 2.5 K on global average over the twentieth century and ERA-20CM by about 2 K. Correlations between the reanalyses and HadEX2 are again moderate to high (0.64 for 20CR and 0.81 for ERA-20C) but considerably lower between the AMIP ensemble and HadEX2. As with warm extremes and TN10p, local correlations between the reanalyses and gridded observations are highest (up to 0.8) over Europe and North America (not shown). The AMIP ensemble, however, hardly shows any significant ($p \leq 0.05$) local correlations with the other data sets.

Again, broadly similar results are found for other indices that measure aspects of cold extremes (shown as supporting information figures), such as the frequency of cold days (TX10p, changes are similar to TN10p), the coldest day of the year (TXn, changes are similar to TNn), or the frequency of frost days (FD).

3.3. Wet Extremes

The number of heavy precipitation days (R10mm) in gridded observations shows some increases in northern Europe and central Eurasia over the past century and also in eastern North America and the east coast of South America since the midtwentieth century (Figure 7). Increases in similar regions are also seen in the century-long reanalyses, but these data sets having complete spatial coverage show the strongest changes in the tropics. Both 20CR and ERA-20C show strong increases of about 10 days over the northern part of South America, but change patterns over Africa are partly different between the two reanalyses. The 20CR indicates a change toward less frequent heavy precipitation days in western central Africa over the past century, while ERA-20C shows slight increases in this region but with a drying in the southeast of Africa. After the midtwentieth century both reanalyses show increases in the western part of tropical Africa and decreases in the eastern part but with different magnitudes. All data sets show changes toward more extreme precipitation over parts of tropical Southeast Asia including Indonesia. Change patterns in the ERA-20CM ensemble appear much smoother than in the other data sets that make use of land-based observations, and local changes in the tropics are mostly of smaller magnitude than in the two reanalysis data sets. ERA-20CM shows

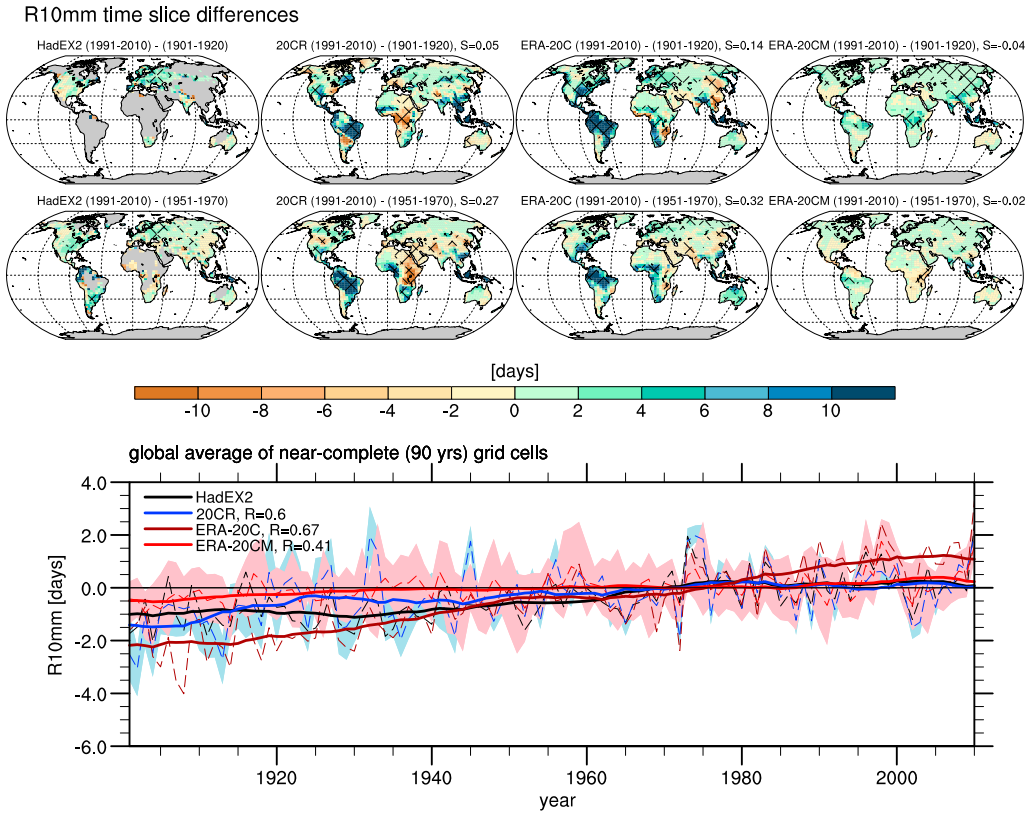

Figure 7. Regional and global changes in heavy precipitation days (R10mm). For detailed description of the plot see Figure 1 caption.

some significant increases of about 2 to 4 days over the past century in northern Eurasia, South Asia, and tropical parts of Africa. In general, spatial correlations between reanalyses/model simulations and gridded observations are low (below 0.3).

Related to the finding that the frequency of precipitation extremes has increased in more regions than it has decreased [e.g., *Donat et al.*, 2013b; *Westra et al.*, 2013], on average all data sets show R10mm increases over the past century. However, the magnitude of increase is very different, ranging from about 1 day on global average in ERA-20CM to more than 3 days in ERA-20C. There are moderate correlations of about 0.65 between the detrended global time series from the reanalyses and gridded observations. Even the AMIP-style simulations still show a significant correlation with the HadEX2 time series of about 0.4. Moderate correlations of up to 0.6 to 0.7 between the gridded observations and reanalyses are also found locally over parts of North America, Europe, and Australia (Figure 8).

All precipitation indices within any one data set show similar spatial patterns of change in that areas of increase and decrease are very similar; however, comparing across different data sets, there is considerable disagreement of change patterns. Therefore, the precipitation amount that falls on very wet days (R95p, Figure 9) or the annual maximum 1 day precipitation total (Rx1day, supporting information Figure S13) shows increases in similar regions where also R10mm increases were found, and the same is true for decreases, respectively. Consequently, the global averages from all data sets show increases also for these extreme precipitation indices. And as with R10mm, these increases are strongest for ERA-20C and weakest for ERA-20CM.

The temporal correlations between data sets become smaller for the indices that represent rarer and more extreme precipitation measures. R95p still shows global average time series correlations of about 0.5 between the reanalyses and gridded observations (Figure 9), while for Rx1day these correlations are only about 0.4 (supporting information Figure S13). These lower correlations are expected given the larger noise

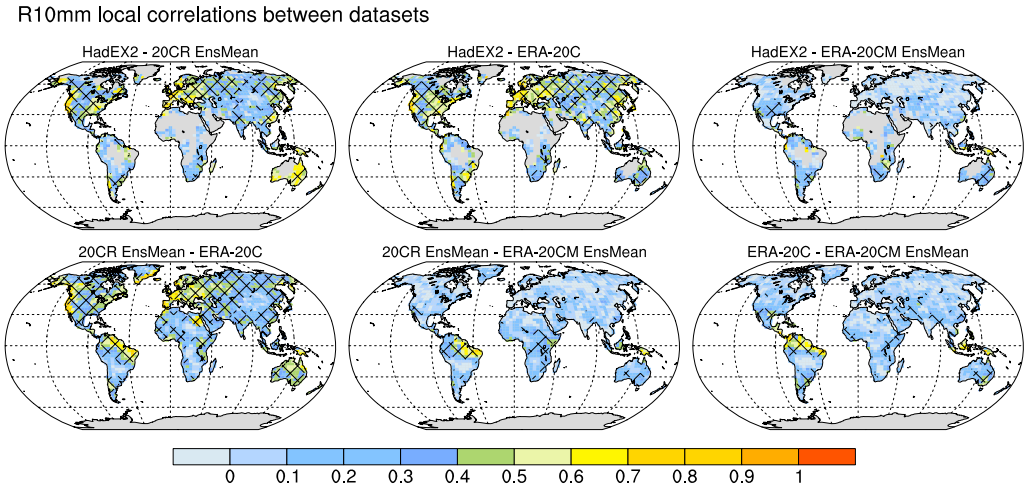

Figure 8. Maps of local correlations between the different data sets for R10mm. The time series from each grid cell were detrended prior to calculating correlations. Hatching indicates where correlations are significant ($p \leq 0.05$).

related to indices that consider only one or few days per year. The correlations of ERA-20CM with HadEX2 are generally close to zero for these more extreme precipitation indices. This means that these more extreme precipitation indices cannot be explained by SST and sea ice variability on an interannual time scale, either on global average (e.g., Figure 9) or locally (supporting information Figure S14).

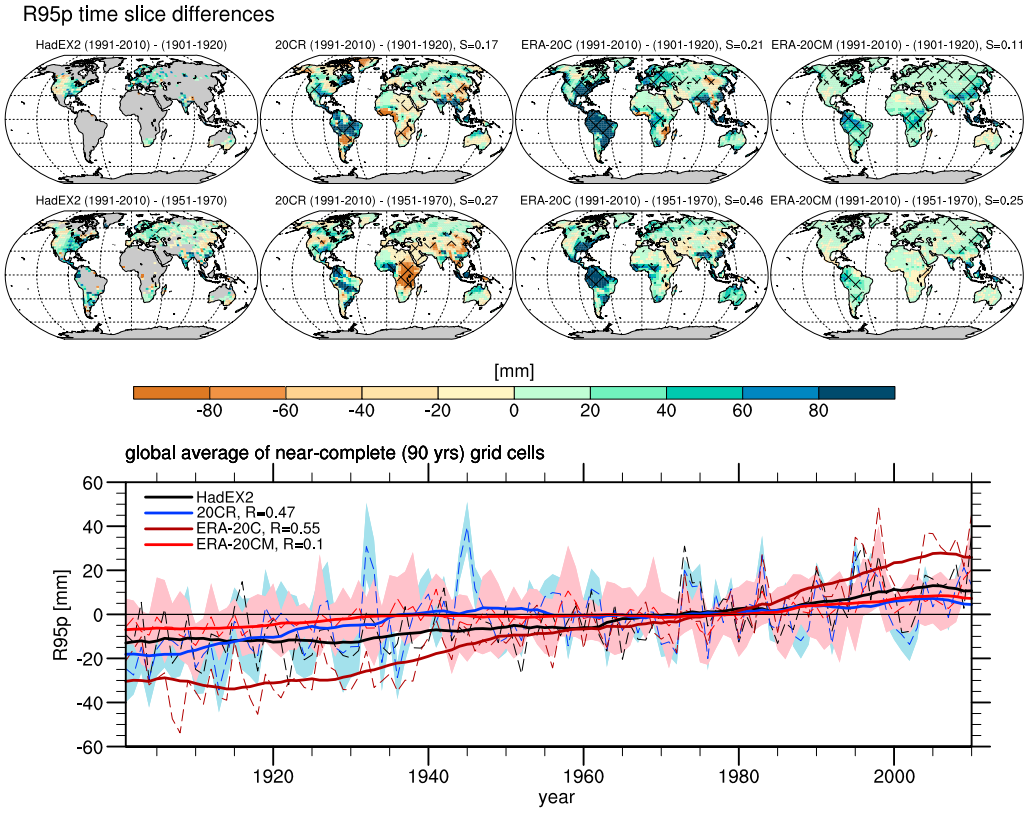

Figure 9. Regional and global changes in precipitation from very wet days (R95p). For detailed description of the plot see Figure 1 caption.

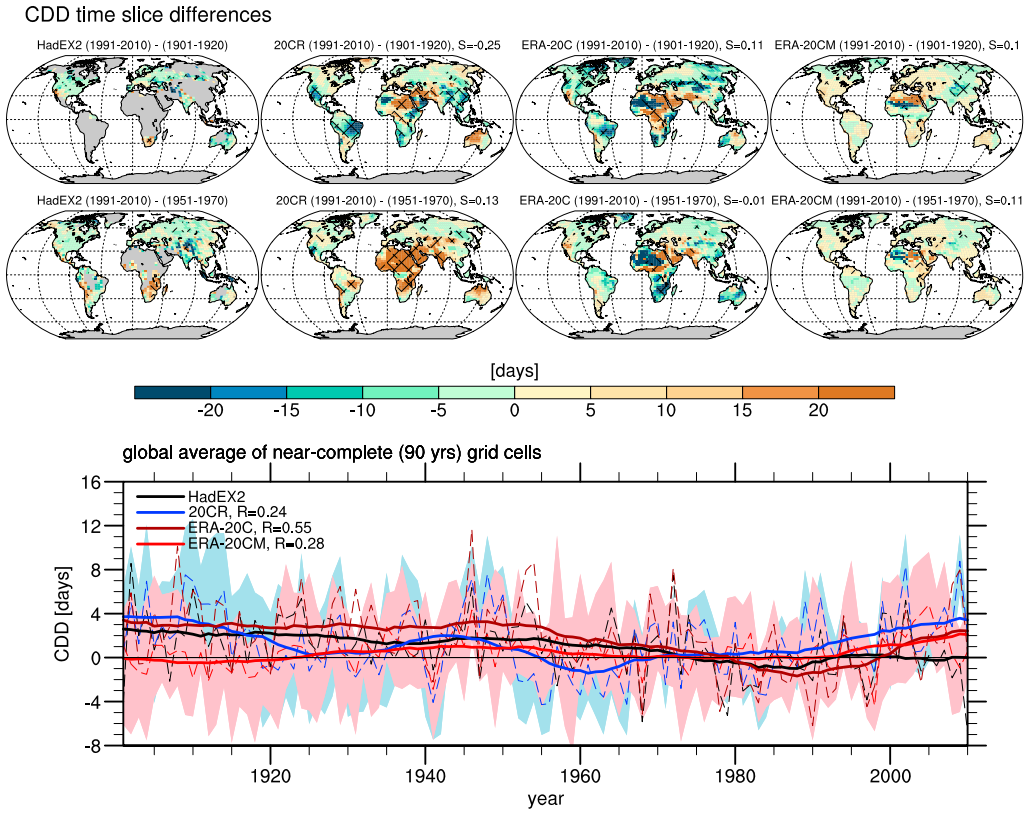

Figure 10. Regional and global changes in consecutive dry days (CDD). For detailed description of the plot see Figure 1 caption.

3.4. Dry Extremes

Only one of the ETCCDI indices, Consecutive Dry Days (CDD), is available as a measure of dry extremes. While on global average no significant changes are apparent (Figure 10), gridded station observations (HadEX2) show CDD increases in southern Africa and along the west coasts of North and South America. Largely different spatial patterns of change are found between the two reanalysis products. The 20CR also shows a strong CDD increase in southern Africa (and most of the rest of the continent where HadEX2 does not have coverage) since the middle of the twentieth century, but ERA-20C mostly shows CDD decreases in these regions. Conversely, ERA-20C reproduces the observed CDD increases along the west coast of the Americas, whereas 20CR shows CDD decreases in some of these regions. There is virtually no spatial correlation between the change patterns from the different data sets.

The global average time series show correlations between 0.3 and 0.5, and moderately high ($R > 0.5$) local correlations between gridded observations and reanalyses are mainly found over Europe (not shown as maps look very similar to other precipitation indices, e.g., Figure 8). These regions, however, generally do not show any significant changes over the past century. Based on the large disagreement in regional changes between the different data sets, we can only have little to no confidence in local CDD changes over the past century, and none of the data sets actually indicate changes on the global scale.

4. Summary and Discussion

We analyze long-term changes and interannual variability of temperature and precipitation extremes across several observations-based data sets that cover the entire twentieth century. These data sets have different levels of observational constraints and include gridded station observations, reanalyses driven by variables related to large-scale atmospheric flow, SSTs and sea ice, and atmospheric model simulations only driven by monthly SSTs and SICs.

All data sets show that warm extremes have increased in frequency and intensity, while cold extremes have decreased in frequency and intensity (consistent with increased temperature values) in most regions of the globe throughout the past century. For most indices warming trends, on average, are strongest for the most recent decades after about 1980. However, spatial details of change differ between data sets. In particular, the two reanalyses, 20CR and ERA-20C, for some indices show relatively warm conditions in the first half of the twentieth century in some regions, mainly in high northern latitudes. The ERA-20CM runs, in contrast, show continuous warming over all land areas for all indices during the twentieth century. This suggests that the warm biases in high northern latitudes in the reanalyses during earlier decades, which also have a clear signature in the global average time series, may be related to the assimilation of pressure observations.

Precipitation indices on average show changes toward more frequent and more intense precipitation extremes, related to the finding that the extremes are increasing in more regions of the globe than they are decreasing. The spatial distribution of increasing and decreasing changes is more heterogeneous than with extreme temperature changes. In particular, there is little agreement between the different data sets in the spatial patterns of change. The only regions where all data sets seem to agree in showing changes toward more extreme precipitation over the past century are northern Europe and central parts of Eurasia. The analyses of long-term changes are, however, limited by sparse observational coverage in the early decades of the twentieth century. The two reanalyses and ERA-20CM also show more frequent and more intense extreme precipitation relative to the early twentieth century over the Amazon region and parts of tropical Southeast Asia including Indonesia. However, no long-term precipitation observations are available for these regions. The one index representing dry extremes does not indicate global changes, and regional changes generally do not agree across the different data sets.

These analyses show that we have relatively high confidence in global-scale changes toward warmer temperature extremes and more intense precipitation extremes, but the time series become increasingly noisy (and changes are often less significant) at regional scales. This effect was shown to be related to internal climate variability in climate model simulations [*Deser et al*., 2012; *Donat*, 2013; *Fischer et al*., 2013]. Our results indicate that this feature of high confidence in global changes but stronger noise at regional scales is also found in observations.

Comparing the ensemble of AMIP runs with data sets that make use of land-based observations, we find moderate correlations of global land average time series for temperature extremes. These results suggest that ocean temperatures may explain about 25% to 50% of the observed variability of warm and cold extremes, at least using this specific atmospheric model from ECMWF. Local correlations, however, are mostly only significant ($p \leq 0.05$) in low latitudes, suggesting that the effect of SSTs on temperature extremes in the extratropics is generally small. For moderate precipitation extremes that, on average, occur on several days per year, the model simulations suggest that ocean temperatures and sea ice may explain up to 15% of the interannual variability (correlation about 0.4) in the global average time series. For more extreme precipitation extremes, however, no significant relationship regarding year-to-year variability could be found. It is worth noting that the ERA-20C values often lie outside the range of the ERA-20CM ensemble, indicating that the atmosphere-only model runs (i.e., without assimilating variables related to the atmospheric circulation) may not capture the specific assimilated weather situations.

For most indices, there is reasonably good agreement in the global average time series from the different data sets over the past 60–70 years, back to the midtwentieth century. But there are larger differences between data sets in the earlier decades of the twentieth century. However, results also show higher correlations of the (detrended) local time series over regions that have good observational coverage, such as North America, Europe, parts of Asia, and Australia. Comparing the regional average time series shows that there is also less spread between the different data sets during the first half of the twentieth century for these regions as compared to the global average, for example. This shows that there is also better agreement between data sets regarding century-scale changes in extremes for these regions with good observational coverage.

However, not all regions have good observational coverage. Reanalysis products, on the other hand, provide complete coverage in space and time. This makes reanalyses interesting candidates as substitutes for the real-world observations for regions where no in situ observations of temperature and precipitation extremes are available. It is, however, questionable whether it is justifiable to use reanalyses to fill in information for areas without reliable observations.

Our results show that the reanalyses show generally good agreement with each other and with gridded observations in regions where good observational networks are in place. They show, however, some different patterns of change in regions without good observational coverage. Even over the past three decades, different modern reanalyses that all also assimilate satellite observations show different change patterns, from each other and compared to observations, for precipitation over Africa [*Schmocker et al*., 2015]. Therefore, the analyzed changes in these regions seem to depend on the specific modeling and data assimilation systems used, and the contrary results show that we are not able to make reliable estimates of change in data-sparse areas. This shows that more long-term, high-quality observations are needed from areas with poor coverage, in particular Africa, South America, and parts of Southeast Asia, to increase our confidence in the changes for those regions, and also in long-term global changes.

It should also be noted that gridded observations themselves may be subject to a number of uncertainties related to quality of the in situ data, the station network, and the gridding methods used [*Hofstra et al*., 2008; *Dunn et al*., 2014; *Gervais et al*., 2014]. This means that HadEX2 should not be considered as "truth" but rather as one possible realization. Therefore, the specific quantifications of interdata set agreement presented here may slightly differ for other methodological or structural choices in the data set production. Previous attempts to quantify the uncertainties related to this data set suggest, however, that global results are relatively robust, whereas uncertainties are largest at the grid box level [*Dunn et al*., 2014].

5. Conclusions

Based on generally good agreement between the different data sets, our results confirm that there is high confidence that temperature extremes have seen warming on a global scale since the midtwentieth century [*Hartmann et al*., 2013]. Due to the larger differences between data sets for the earlier twentieth century, our confidence in earlier changes is lower. However, for most indices, and in particular for all indices of cold extremes, the most recent decades show warmer conditions than in any previous decade on record, in all data sets globally and in most regions. Therefore, we are also confident to state that cold extremes on a global scale have become warmer and less frequent over the past century.

All data sets agree that precipitation extremes have intensified on global average over the past century, suggesting that we can robustly conclude increasing precipitation extremes on the global scale over the past century. However, there is mostly little agreement between the different data sets regarding specific regions where precipitation extremes have intensified or weakened, so regional extreme precipitation changes over the past century are generally not robust, apart from maybe northern Europe and some central parts of midlatitudinal Eurasia where all data sets show tendencies toward more extreme precipitation over the past century. Based on the analyses presented in this study, possible regional changes in dry extremes over the past century are generally not robust, and none of the data sets indicates global changes in dry extremes.

More long-term and high-quality station observations of daily scale temperature and precipitation extremes that extend back into the early twentieth century are needed from larger parts of Africa, South America, and Southeast Asia for assessing century-long changes in these regions with some confidence.

Comparison between gridded observations and model simulations driven by observed SST and sea ice fields suggests that up to about 50% of the interannual variability of temperature extremes, on global average, can be explained by variations in these ocean variables. Regionally, these relationships are strongest in low latitudes. To a lesser extent, moderate precipitation extremes also appear to be related to SST and sea ice variability, but no significant relationship was found for the more extreme precipitation extremes. These estimates of explained variability, however, are obtained from simulations with only one atmospheric model and might be subject to biases in that specific model. Similar analyses should be performed for several different models forced by ocean surface variability, as coordinated in the context of AMIP [*Gates*, 1992] or the Climate of the Twentieth Century (Plus) projects [*Scaife et al*., 2008; *Folland et al*., 2014], for example. This will enable a more systematic and more comprehensive understanding of interannual variability in extremes.

Acknowledgments
This study was supported through Australian Research Council grants CE110001028 and DE150100456. L.V.A. is also supported by ARC grant DP160103439. We are grateful to Gil Compo for providing daily data from all 20CR ensemble members to calculate the extremes indices. Support for the 20CR Project is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. All data used in this study are freely available. HadEX2 can be downloaded from www.climdex.org, 20CR data are available from http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html, and ERA-20C/ERA-20CM can be accessed at the ECMWF data server via http://apps.ecmwf.int/datasets/data/era20c-daily/levtype=sfc/type=an/ and http://apps.ecmwf.int/datasets/data/era20cm-daily/levtype=sfc/.

References

Alexander, L. V., et al. (2006), Global observed changes in daily climate extremes of temperature and precipitation, *J. Geophys. Res.*, *111*, D05109, doi:10.1029/2005JD006290.

Avila, F. B., S. Dong, K. P. Menang, J. Rajczak, M. Renom, M. G. Donat, and L. V. Alexander (2015), Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia, *Weather Clim. Extrem.*, *9*, 6–16, doi:10.1016/j.wace.2015.06.003.

Bengtsson, L., V. A. Semenov, and O. M. Johannessen (2004), The early twentieth-century warming in the Arctic—A possible mechanism, *J. Clim.*, *17*(20), 4045–4057, doi:10.1175/1520-0442(2004)017<4045:TETWIT>2.0.CO;2.

Compo, G. P., et al. (2011), The Twentieth Century Reanalysis Project, *Q. J. R. Meteorol. Soc.*, *137*(654), 1–28, doi:10.1002/qj.776.

Compo, G. P., P. D. Sardeshmukh, J. S. Whitaker, P. Brohan, P. D. Jones, and C. McColl (2013), Independent confirmation of global land warming without the use of station temperatures, *Geophys. Res. Lett.*, *40*, 3170–3174, doi:10.1002/grl.50425.

Cornes, R. C., and P. D. Jones (2013), How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe?, *J. Geophys. Res. Atmos.*, *118*, 10,262–10,276, doi:10.1002/jgrd.50799.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips (2012), Communication of the role of natural variability in future North American climate, *Nat. Clim. Chang.*, *2*(11), 775–779, doi:10.1038/nclimate1562.

Donat, M. G. (2013), Projection and prediction: Local noise and global confidence, *Nat. Clim. Chang.*, *3*(12), 1018–1019.

Donat, M. G., D. Renggli, S. Wild, L. V. Alexander, G. C. Leckebusch, and U. Ulbrich (2011), Reanalysis suggests long-term upward trends in European storminess since 1871, *Geophys. Res. Lett.*, *38*, L14703, doi:10.1029/2011GL047995.

Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar (2013a), Global land-based datasets for monitoring climatic extremes, *Bull. Am. Meteorol. Soc.*, *94*(7), 997–1006, doi:10.1175/BAMS-D-12-00109.1.

Donat, M. G., et al. (2013b), Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, *J. Geophys. Res. Atmos.*, *118*, 2098–2118, doi:10.1002/jgrd.50150.

Donat, M. G., J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers (2014), Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets, *J. Clim.*, *27*(13), 5019–5035.

Donat, M. G., A. D. King, J. T. Overpeck, L. V. Alexander, I. Durre, and D. J. Karoly (2016), Extraordinary heat during the 1930s US Dust Bowl and associated large-scale conditions, *Clim. Dyn.*, *46*(1–2), 413–426, doi:10.1007/s00382-015-2590-5.

Drinkwater, K. F. (2006), The regime shift of the 1920s and 1930s in the North Atlantic, *Prog. Oceanogr.*, *68*(2–4), 134–151, doi:10.1016/j.pocean.2006.02.011.

Dunn, R. J. H., M. G. Donat, and L. V. Alexander (2014), Investigating uncertainties in global gridded datasets of climate extremes, *Clim. Past*, *10*(6), 2171–2199, doi:10.5194/cp-10-2171-2014.

Fischer, E. M., U. Beyerle, and R. Knutti (2013), Robust spatially aggregated projections of climate extremes, *Nat. Clim. Chang.*, *3*(12), 1033–1038, doi:10.1038/nclimate2051.

Folland, C., D. Stone, C. Frederiksen, D. Karoly, and J. Kinter (2014), The international CLIVAR Climate of the 20th Century Plus (C20C+) project: Report of the sixth workshop, *CLIVAR Exch.*, *19*, 57–59.

Gates, W. L. (1992), AMIP: The Atmospheric Model Intercomparison Project, *Bull. Am. Meteorol. Soc.*, *73*(12), 1962–1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.

Gervais, M., L. B. Tremblay, J. R. Gyakum, and E. Atallah (2014), Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods, *J. Clim.*, *27*(14), 5201–5218, doi:10.1175/JCLI-D-13-00319.1.

Hartmann, D. J., et al. (2013), Observations: Atmosphere and surface, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by P. M. Stocker et al., pp. 159–254, Cambridge Univ. Press, Cambridge.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee (2015), ERA-20CM: A twentieth-century atmospheric model ensemble, *Q. J. R. Meteorol. Soc.*, *141*, 2350–2375, doi:10.1002/qj.2528.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, *113*, D21110, doi:10.1029/2008JD010100.

Intergovernmental Panel on Climate Change (2012), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, edited by P. M. Field et al., Cambridge Univ. Press, Cambridge, U. K.

King, A. D., L. V. Alexander, and M. G. Donat (2013), Asymmetry in the response of eastern Australia extreme rainfall to low-frequency Pacific variability, *Geophys. Res. Lett.*, *40*, 2271–2277, doi:10.1002/grl.50427.

Krueger, O., F. Schenk, F. Feser, and R. Weisse (2013), Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations, *J. Clim.*, *26*(3), 868–874, doi:10.1175/JCLI-D-12-00309.1.

Misra, V., J.-P. Michael, R. Boyles, E. P. Chassignet, M. Griffin, and J. J. O'Brien (2012), Reconciling the spatial distribution of the surface temperature trends in the southeastern United States, *J. Clim.*, *25*, 3610–3618, doi:10.1175/JCLI-D-11-00170.1.

Pepler, A. S., J. Fong, and L. V. Alexander (2016), Australian east coast mid-latitude cyclones in the 20th Century Reanalysis ensemble, *Int. J. Climatol.*, doi:10.1002/joc.4812.

Poli, P., et al. (2016), ERA-20C: An atmospheric reanalysis of the 20th century, *J. Clim.*, JCLI–D–15–0556.1, doi:10.1175/JCLI-D-15-0556.1.

Portmann, R. W., S. Solomon, and G. C. Hegerl (2009), Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States, *Proc. Natl. Acad. Sci. U.S.A.*, *106*(18), 7324–9, doi:10.1073/pnas.0808533106.

Salinger, M. J. (2005), Climate variability and change: Past, present and future—An overview, *Clim. Change*, *70*(1–2), 9–29, doi:10.1007/s10584-005-5936-x.

Scaife, A. A., et al. (2008), The CLIVAR C20C project: Selected twentieth century climate events, *Clim. Dyn.*, *33*(5), 603–614, doi:10.1007/s00382-008-0451-1.

Schmocker, J., H. P. Liniger, J. N. Ngeru, Y. Brugnara, R. Auchmann, and S. Brönnimann (2015), Trends in mean and extreme precipitation in the Mount Kenya region from observations and reanalyses, *Int. J. Climatol.*, *36*, 1500–1514, doi:10.1002/joc.4438.

Seneviratne, S. I., M. G. Donat, B. Mueller, and L. V. Alexander (2014), No pause in the increase of hot temperature extremes, *Nat. Clim. Chang.*, *4*(3), 161–163.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh (2013), Climate extremes indices in the CMIP5 multimodel ensemble: 1. Model evaluation in the present climate, *J. Geophys. Res. Atmos.*, *118*, 1716–1733, doi:10.1002/jgrd.50203.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, D. Bronaugh, and M. G. Donat (2014), Evaluating model-simulated variability in temperature extremes using modified percentile indices, *Int. J. Climatol.*, *34*(11), 3304–3311, doi:10.1002/joc.3899.

Skansi, M. de l. M., et al. (2013), Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America, *Glob. Planet. Change*, *100*, 295–307, doi:10.1016/j.gloplacha.2012.11.004.

Thorne, P. W., and R. S. Vose (2010), Reanalyses suitable for characterizing long-term trends, *Bull. Am. Meteorol. Soc.*, *91*(3), 353–361, doi:10.1175/2009BAMS2858.1.

Titchner, H. A., and N. A. Rayner (2014), The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, *J. Geophys. Res. Atmos.*, *119*, 2864–2889, doi:10.1002/2013JD020316.

Wang, X. L., Y. Feng, G. P. Compo, F. W. Zwiers, R. J. Allan, V. R. Swail, and P. D. Sardeshmukh (2013), Is the storminess in the Twentieth Century Reanalysis really inconsistent with observations? A reply to the comment by Krueger et al. (2013b), *Clim. Dyn.*, *42*(3–4), 1113–1125, doi:10.1007/s00382-013-1828-3.

Westra, S., L. V. Alexander, and F. W. Zwiers (2013), Global increasing trends in annual maximum daily precipitation, *J. Clim.*, *26*, 3904–3918, doi:10.1175/JCLI-D-12-00502.1.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers (2011), Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wiley Interdiscip. Rev. Clim. Chang.*, *2*(6), 851–870, doi:10.1002/wcc.147.